CLAIMS

1.   In a file server having a storage operating system, a
method for managing storage of data in a plurality of
storage devices, each storage device comprising a plurality
of blocks for storing data, the method comprising the steps
of:

generating block layout information in a file system
layer of the storage operating system by determining which
blocks within the plurality of blocks are allocated for
storing data and which are unallocated;

transferring the block layout information from the file
system layer to a RAID layer of the storage operating
system; and

responsive to the block layout information, the RAID
layer controlling the execution of I/O operations by
identifying blocks within the plurality of blocks for use by
the I/O operations so as to substantially maximize chain
lengths of reads for calculation of parity and of writes for
data write operations thereof while substantially minimizing
cost of calculation of parity,

wherein the block identification comprises the steps of

examining all the I/O operations,

selecting a method for parity calculation which
substantially minimizes cost of the parity calculation
for the I/O operations, and

responsive to the block layout information and the
parity calculation method selection, identifying the
blocks within the plurality of blocks for use by the
I/O operations.

1   2. A method for managing storage of data in a plurality of
2   storage devices, each comprising a plurality of storage
3   blocks, the method comprising the steps of:

4   generating block layout information; and

5   in response to the block layout information,
6   controlling the execution of an I/O operation by
7   identifying storage blocks for use by the I/O operation
8   so as to substantially minimize cost of calculation of
9   error correction parameters across a stripe.

1   3. The method of claim 2 wherein the calculation of error
2   correction parameters comprises the calculation of
3   parity.

1   4. The method of claim 3 wherein the calculation of parity
2   comprises selecting a parity calculation operation from a
3   group consisting of a subtraction method and a parity re-
4   calculation method.

1   5. The method of claim 2 wherein the identification of
2   storage blocks for use in the I/O operation substantially
3   maximizes the chain length by substantially maximizing
4   the number of blocks having a contiguous physical layout
5   on a storage device.

1   6. The method of claim 2, further comprising the step of
2   identifying storage blocks for use in the I/O operation
3   so as to substantially maximize the chain length by
4   substantially maximizing the number of blocks having
5   sequential VBN's associated with the storage blocks.

1   7. The method of claim 2, further comprising the step of
2   identifying storage blocks for use in the I/O operation
3   so as to substantially maximize the chain length by
4   substantially maximizing the locality of the blocks of a
5   storage device.

1   8. The method of claim 2 wherein the execution controlling
2   step comprises the steps of:

3   examining blocks to which data is to be written prior
4   to write operations;

5   selecting one of a plurality of parity calculation
6   methodologies including a first methodology comprising
7   minimizing the number of blocks read, and a second
8   methodology comprising maximizing chain lengths of
9   blocks read for the parity calculation.

1   9. The method of claim 8, wherein the execution controlling
2   step further comprises the steps of:

3   implementing the selection responsive to the block
4   layout information; and

5   wherein, if the selection constitutes substantially
6   minimizing the number of read blocks,

7   determining on a stripe-by-stripe basis whether to
8   calculate parity based on a subtraction method or a
9   recalculation method,

10   performing any appropriate read operations to
11   support the method selected, and

12   calculating parity responsive to the read blocks and
13   the data to be written; and

14   wherein, if the selection constitutes substantially
15   maximizing chain lengths of blocks read,

16   deciding which storage blocks to read to
17   substantially maximize chain length while substantially
18   minimizing the number of storage blocks read to support
19   either a subtraction method or a recalculation method,

20   performing read operations on the blocks to be read,
21   and

| 22 | calculating parity responsive to the read blocks and |
| 23 | the data to be written. |

1 10. The method of claim 2, wherein the identification of
2   storage blocks is based at least in part on an available
3   resource.

1 11. The method of claim 2 further comprising transmitting
2   the block layout information from a file system layer to
3   a RAID layer.

1 12. The method of claim 2 wherein the generating step
2   further comprises:

3   making a first determination as to whether a storage
4   block is unallocated;

5   making a second determination as to a current
6   implementation of the plurality of storage devices; and

7   generating the block layout information based at least
8   in part on the first and the second determinations.

1 13. The method of claim 2, wherein the I/O operation is one
2   of a plurality of I/O operations and one of the plurality
3   of I/O operations is a read operation.

1 14. The method of claim 2, wherein the chain length is a
2   chain length of a read operation for calculation of
3   parity.

1 15. The method of claim 2, wherein the chain length is a
2   chain length for a write operation for the data.

1 16. A method for managing storage of data in a storage
2   system comprising a plurality of storage devices each
3   comprising a plurality of storage blocks, the method
4   comprising writing data to predetermined storage blocks
5   across a plurality of stripes and to predetermined
6   storage blocks within each storage device so as to

7   substantially maximize chain length of storage blocks

8   within each storage device while substantially minimizing

9   cost of calculation of error correction parameters across

10   each stripe of the plurality of stripes.

1   17. A method for controlling storage of data in a plurality

2   of storage devices each comprising storage blocks, the

3   method comprising:

4   buffering a plurality of write requests associated

5   with data blocks for a single write transaction;

6   defining a group of storage blocks, the group

7   comprising a plurality of storage blocks in each of the

8   plurality of storage devices; and

9   associating each data block with a respective one of

10   the storage blocks, for transmitting the association to a

11   storage device manager for processing of the single write

12   transaction.

1   18. The method of claim 17, further comprising receiving by

2   a file system the plurality of write requests, and

3   transmitting from the file system to the storage device

4   manager the association for processing of the single

5   write transaction, wherein the storage device manager

6   comprises a RAID layer.

1   19. The method of claim 17, wherein associating comprises

2   associating each data block of at least one of the write

3   requests with storage blocks of only one of the plurality

4   of storage devices.

1   20. The method of claim 17, wherein defining the group of

2   storage blocks comprises receiving by the file system

3   from the storage device manager configuration information

4   of the plurality of storage devices.

1   21. The method of claim 20, wherein the configuration

2   information comprises disk topology information.

1   22. The method of claim 17, wherein the group of storage

2   blocks comprises more than one stripe in a RAID group.

1   23. The method of claim 22, wherein the group of storage

2   blocks comprises more than one contiguous stripe.

1   24. The method of claim 23, further comprising transmitting

2   the association and the data to the plurality of storage

3   devices to store each data block at the respective one of

4   the storage blocks of the more than one contiguous

5   stripe.

1   25. The method of claim 22, wherein the plurality of storage

2   devices comprises more than one RAID group, and defining

3   the group of storage blocks comprises defining the group

4   of storage blocks from storage blocks in a first RAID

5   group.

1   26. The method of claim 25, further comprising defining a

2   second group of storage blocks in a second RAID group for

3   association with data associated with a second plurality

4   of write requests for processing of a second write

5   transaction.

1   27. The method of claim 17, further comprising queuing a

2   plurality of associations by the storage device manager

3   for processing of a plurality of write transactions.

1   28. The method of claim 27, further comprising processing

2   the plurality of write transactions by the storage

3   manager in an order derived from a plurality of

4   priorities associated with the plurality of write

5   transactions.

1   29. The method of claim 17, wherein the group of storage

2   blocks comprises a same quantity of storage blocks in

3   each one of the plurality of storage devices.

1 30. The method of claim 29, wherein each of the plurality of
2   storage blocks comprises contiguous storage blocks.

1 31. The method of claim 17, wherein the group of storage
2   blocks comprises allocated and unallocated storage
3   blocks.

1 32. The method of claim 31, wherein defining the group
2   further comprises selecting an unallocated storage block
3   having a lowest block identifier of any unallocated
4   storage block as a storage block having a lowest storage
5   block identifier of any storage block in the group.

1 33. The method of claim 17, wherein the write requests
2   comprise data to be written.

1 34. A method for storing data blocks, the method comprising:
2   providing a RAID layer in communication with a
3   plurality of storage devices that each comprise a
4   plurality of storage blocks;
5   receiving by the RAID layer a write transaction
6   request that includes an association of each data block
7   with a respective one of a group of storage blocks, the
8   group comprising a plurality of storage blocks in each of
9   the plurality of storage devices; and
10   storing the data blocks by the RAID layer in the group
11   of storage blocks according to the association.

1 35. The method of claim 34, wherein the RAID layer comprises
2   a second plurality of storage devices that each
3   comprise a plurality of storage blocks that mirror
4   storage blocks of the plurality of storage devices, and
5   further comprising storing the data in a group of
6   storage blocks of the second plurality of storage
7   devices.

1   36. The method of claim 35, further comprising performing a
2   parity determination for one of the groups of storage
3   blocks prior to storing, and sharing a result of the
4   parity determination with the other of the groups of
5   storage blocks.

1   37. The method of claim 36, further comprising selecting one
2   of the groups of storage blocks for performing the
3   parity determination.

4   38. The method of claim 34, further comprising monitoring by
5   the storage device manager of configuration information
6   of the plurality of storage devices.

1   39. A storage system comprising:

2   a plurality of storage devices each comprising:

3   a plurality of storage blocks; and

4   a storage manager in communication with the plurality
5   of storage devices, the storage manager writing data to
6   predetermined storage blocks across a plurality of
7   stripes and to predetermined storage blocks within each
8   storage device so as to substantially maximize chain
9   length of storage blocks within a storage device while
10   substantially minimizing the calculation of error
11   correction parameters across each stripe of the plurality
12   of stripes.

1   40. A system for managing the storage of data, the system
2   comprising:

3   a plurality of storage devices each having a plurality
4   of storage blocks;

5   a storage device manager in communication with the
6   plurality of storage blocks;

7   a block layout information generator in communication

8   with the storage device manager and the plurality of

9   storage blocks; and

10   an error correction parameter calculator in

11   communication with the plurality of storage blocks and

12   the storage device manager,

13   wherein the storage device manager, in response to the

14   block layout information from the block layout

15   information generator, controls the execution of an I/O

16   operation by identifying storage blocks for use by the

17   I/O operation so as to substantially maximize chain

18   length within the storage device while substantially

19   minimizing the calculation by the error correction

20   parameter calculator of error correction parameters

21   across a stripe.

1 41. A method for managing storage of data in storage

2   blocks, the method comprising the steps of:

3   generating block layout information;

4   dynamically determining a first number of error

5   correction calculations;

6   dynamically determining a second number corresponding

7   to a chain length; and

8   in response to the block layout information,

9   controlling the execution of an I/O operation by

10   identifying storage blocks for use by the I/O operation

11   so as to have a chain length of the second number within

12   a storage device while performing the first number of

13   calculations of error correction parameters across a

14   stripe.

1 42. A device for controlling storage of data in a plurality

2   of storage devices each comprising storage blocks, the

3   device comprising:

4   a buffer collecting write requests, each request
5   associated with data blocks;

6   a processor defining a group of storage blocks, the
7   group comprising a plurality of storage blocks in each of
8   the plurality of storage devices, the processor
9   associating each data block with a respective one of the
10   storage blocks for a single write transaction; and

11   a transmitter transmitting the association to a
12   storage device manager for processing of the single write
13   transaction.

1   43. An apparatus for storing data blocks, the apparatus
2   comprising:

3   a plurality of storage devices that each comprise a
4   plurality of storage blocks; and

5   a storage device manager in communication with the
6   plurality of storage devices, and configured to receive a
7   write transaction from a file system, the write
8   transaction comprising an association of each data block
9   to a respective storage block of a group of storage
10   blocks, the group comprising a plurality of storage
11   blocks in each of the plurality of storage devices.

1   44. The apparatus of claim 43, wherein the storage device
2   manager causes a parity determination for the group of
3   storage blocks prior to storing the data blocks in the
4   group of storage blocks.

2316297_1

52